

Abstract Title Page
Not included in page count.

Title: The relation of the response distribution to self-report questionnaire items and cognitive ability among school-aged children

Author(s):

Laura M. Stapleton, University of Maryland, Baltimore County

Laura S. Reiner, University of Maryland, Baltimore County

Anu J. Aluvathingal, University of Maryland, Baltimore County

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

As part of experiments in educational contexts, self-report surveys of children are sometimes used, both to measure outcomes and to examine program implementation fidelity. For example, in one of the Institute for Education Sciences (IES)-funded studies (award # R324G06039), self-report by 2nd graders is planned to be used to assess the effects of instruction on attitudes and engagement as well as on knowledge of reading (Institute for Education Sciences, 2009a). In another IES-funded study (award number unknown, principal investigator is Nancy Romance) self-report by 3rd through 7th graders was reported to be used to determine the use of reading comprehension strategies and also to evaluate students' attitudes and self-confidence in reading (Institute for Education Sciences, 2009b). In addition, administrators increasingly use surveys to obtain climate information from students. In a recent nationally-representative survey of elementary school principals, approximately 55 percent of elementary schools reported conducting surveys of students on a regular basis: 80% of schools survey 7th/8th graders, about half survey 3rd/4th/5th graders, and 20-25% survey K-2nd graders (Stapleton, Cafarelli, Almario & Ching, in press). In nearly three-fourths of the schools, responses are compared across grades, even though evidence suggests that children at various developmental levels do not use the same survey response process.

The cognitive response model is described as having four components: comprehension (understanding the question), retrieval (gathering information from memory), judgment (evaluating retrieved information), and communication (translating the information into a response) (Tourangeau, Rips, & Rasinski, 2000). At each of these steps there is room for measurement error no matter the age of the respondent. This study focuses on the last step of the process, the communication of response. Can children appropriately map their retrieved, mental, response to a question onto the response scale provided by a researcher? The design of the response option set has been found to affect measurement quality. The optimal number of response options for Likert-type items depends on the context of the question; most researchers suggest between five and nine options for the adult population, with fewer options for less educated respondents. Also, for frequency scales, when appropriate, concrete response options (such as *everyday* and *once or twice a week*) lead to more highly reliable measurement as opposed to vague quantifiers (such as *always*, *most of the time*, and *rarely*; Dillman, 2000).

Children's response quality has been found to be function of their cognitive and social-cognitive development (Stone & Lemanek, 1990). Several developmental stages have been identified: very young or preoperational (three to six/seven years of age), concrete operational (seven/eight to 11/12 years of age), and adolescents (12 years and older). These age groupings translate to school grades of kindergarten to 1st/2nd grade, 3rd grade to 6th/7th grade, and 7th grade and above. Children below the age of seven "do not have sufficient cognitive skills to be effectively...questioned" (de Leeuw, 2005, p.1) therefore researchers encourage qualitative, open-ended interviews with very young children. Children in the early stages of development tend to be literal, interpreting words in unanticipated ways (Borgers, de Leeuw, & Hox, 2000). Entrusting the children to read or listen to questions and understand the intended content without

some probing for comprehension is problematic, and yet Stapleton et al. (in press) found that 20-25% of schools reported surveying kindergarten through 2nd graders. Whereas Stanford, Chambers and Craig (2006) found that young children (ages 3 to 6) could accurately use a self-report scale for pain in response to constructed vignettes, Rebok et al. (2001), in their cognitive interviewing studies of children aged 5 to 11, found that 5 year old children did not sufficiently understand written questions to report on their own health and while 6- and 7-year-old children understood the question, they responded at the extremes of a response scale of graduated circles. These 6- and 7-year-olds gave extreme answers to 79% and 61% of questions, respectively, at much higher rates than older children (who provided an extreme response to about 50% of the questions). The authors reported that the relation between age and percentage of extreme responses was strong, $r = -.62$.

Older children can still present problems in self-report data collection efforts. Woolley, Bowen, and Bowen (2004) undertook cognitive pretesting with groups of 3rd and 5th graders and found items on scales to be too abstract for the 3rd graders, including statements such as *I feel good about myself* and *I am happy with myself*. The researchers had more success once items were changed to more concrete statements such as *I am smart* and *I am good at art*. Additionally, because of the difficulty in cognitively processing vague quantifiers such as *strongly* and *somewhat*, researchers suggest using simple *yes* and *no* type of responses (Rebok et al., 2001; de Leeuw, Borgers, & Smits, 2004). Other studies have not been as conclusive. Borgers, Hox and Sikkels (2003) found no relation between the use of vague quantifiers and measurement error with this age group. The use of visuals or graphics such as circles growing from small to big or changes in drawn faces that represent levels of happiness, has been found to be successful (Rebok et al., 2001) and this technique was found to be used in some schools (Stapleton et al., in press).

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

This proposed research is part of an on-going line of research of developing questionnaire instruments for use at the elementary school level. Because field trials often use child self-report as outcome measures and sometimes determine implementation fidelity using such measures, evaluation of the validity of the use of such measures with school-aged children is important. Additionally, schools seek to evaluate their learning environments by surveying students, thus it is crucial to determine how to obtain valid measurement of student perceptions. Specifically, this research proposes to answer the following general question:

- Does the cognitive age of the respondent relate to the likelihood of selecting certain response options?

Given that Rebok et al. (2001) found that younger children tended to use extreme values on scales using questions about health, it may be that this same behavior will be found on survey items about school experiences. As a simple visual example, using data from the Progress in International Reading Literacy Study (National Center for Education Statistics, 2006), on questions about school (unrelated to reading specifically), students with lower reading ability tended to use the extreme answers more frequently than peers with higher reading ability for several questions (see Figure 1 for an example).

Setting:

Description of where the research took place.

The research is being undertaken using two sources of secondary data: one collected via a national probability sample and one collected within a health management data collection system in Pennsylvania. Specifically, the first set of data is from the PIRLS, an international survey of 4th grade children and their experiences and ability in reading. The data used in this study only include responses from U.S. students. Because PIRLS is limited to same-age respondents, we also use a clinical sample of children, aged 8 to 13, who are receiving mental health services as part of a health management system within the state of Pennsylvania. Data from respondents older than 13 years old will not be analyzed.

Population / Participants / Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

PIRLS was administered to fourth grade students across the United States, and although ages ranged from 7 to 13 years old with a mean of 10, age was fairly homogenous within the sample (73% of the students were between 9.5 and 10.5 years old). Of the 5,190 students who participated, 50% were female and 50% were male. Of those who indicated their race/ethnicity, 54% were white, 23% were Hispanic, 18% were black, 3% were Asian, 2% were American Indian, and 1% reported multiple races.

There are 968 children who responded to attitude and behavior questions in the health management system. These children are nearly equally distributed across the ages of 8, 9, 10, 11, 12, and 13, with slightly more 10 year olds than others. For those with reported race/ethnicity, 56% are white, 25% are African American, 11% multiracial, 5% are Hispanic, and 3% are Asian American. These children who participate in mental health services range from those receiving out patient services to those who are in foster treatment care.

Intervention / Program / Practice:

Description of the intervention, program or practice, including details of administration and duration.

The study utilizes a secondary data set from a national probability sample and a clinical population and therefore this section is not relevant (see Research Design, below, for a description of the procedures in data collection).

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

This study utilizes a correlational research design; secondary data analysis of survey response data is being undertaken. Our independent variable, developmental age, is not a measure that cannot be manipulated (similar to gender). Although it would be preferable to undertake a study in which we manipulate the feeling of the children (to guarantee the same underlying amount of the construct about which they are reporting) and then determine their use of the response categories, the sample size needed for such an evaluation is prohibitive. We recognize that we

will be unable to indicate a causal relation between cognitive age and response tendency; it is possible that a confound exists.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

The PIRLS study was coordinated by the International Association for the Evaluation of Educational Achievement (IEA). The data analyzed for our study were collected from PIRLS administered in 2006. The sample consists of 5,190 fourth grade students from 183 schools from all across the country. First, schools were randomly selected and then one or two classrooms were randomly selected within each school. The students were assessed on their reading ability, reading achievement, and attitudes because this population represents an important stage in the development of reading. Participation was voluntary and written consent from students' parent or guardian was also obtained. The students were given a reading assessment, and in addition, answered a 24-item questionnaire. Students filled out their assessment and questionnaire in their own classes and in all, PIRLS took approximately 1½ to 2 hours of each student's time. Within the questionnaire, the following items were asked with the possible response options of *agree a lot*, *agree a little*, *disagree a little*, and *disagree a lot*: a) I like being in school, b) I feel safe when I am at school, c) Students in my school show respect to each other, d) Students in my school care about each other, and e) Students in my school help each other with their work. Our analyses specifically examine the response distribution on these five items.

Our second data set is provided by a children's outcomes management center. This center coordinates the mental health outcomes and treatment plans for children being served by a variety of health management organizations and state-run facilities. As part of the data collection and management process, clinicians enter data about the child each quarter, and children and parents are also requested to provide information each quarter on symptoms as well as attitudes. For this study, we use the youth measures obtained at intake, at which time they are asked the frequency of the following items: 1) I attend school regularly, b) I complete school tasks on time, c) I complete homework regularly, d) I pay attention in class, e) I take part in school activities, f) I take part in organized recreational activities, g) I take part in religious activities, h) I have things to do, like hobbies that occupy my time, i) I get along with others my age, j) I have friends that I like to spend time with, and k) I think about how others feel. The children are asked to respond to these items on a five-point frequency scale with the anchors *Never*, *Hardly ever*, *Some of the time*, *Almost always*, and *Always*. The data collection occurs on-line. Parents are requested to have children eight years old and older answer the questions, and they are requested to have the child answer questions by him or herself, with no one else present in the room (other questions on the intake form include information about family functioning, emotional and behavioral symptoms, and treatment progress).

In order to determine whether the cognitive age of respondents is related to the response distribution on the items, ordinal regression will be used, estimating the log of the odds of answering one category over another with cognitive age as a covariate. Each of the response option sets in PIRLS includes four possible responses and thus there are three thresholds to estimate; with the mental health data, there will be four thresholds to estimate. We expect that "younger" respondents will be more likely to answer at the extremes of the response scales and thus, for PIRLS, the first and third thresholds will be closer to the middle of the latent

distribution for younger children; we hypothesize that the 1st threshold will be negatively related to cognitive age and the 3rd threshold will be positively related. For data from the mental health system, we hypothesize that the 1st threshold will be negatively related to cognitive age and the 4th threshold will be positively related. Because the parallel lines assumption, then, is not appropriate in this case, for each outcome variable instead of running one ordinal regression, we will run three separate logistic regressions, controlling type I error with a Bonferroni-type adjustment. The same analyses will be undertaken with the responses to items on both the PIRLS questions and the questions from the mental health management system.

Because the PIRLS data come from a multistage sampling design, the traditional logistic regression standard errors are expected to be negatively biased. Therefore, we will use linearized estimates of the standard errors, accounting for the strata and cluster (school) using PROC SURVEYLOGISTIC in SAS version 9.

All students who respond to the PIRLS questionnaire are close to the same chronological age therefore, when using these data, the reading ability test score will be a proxy measure for the cognitive “age” of the child. Because the entire test was not given to each child, this test score has five plausible values for each child and thus the analyses will be repeated five times and sampling variances of estimates will be determined aggregating across the five analyses. (National Center for Education Statistics, 2006) With the data from the mental health services system, we will use age in years as the measure of cognitive age.

Findings / Results:

Description of main findings with specific details.

Preliminary analyses, using data from PIRLS, indicate that of the five questions that were examined, two fully support the hypothesis (the first and last) and mixed results were found for the remaining three questions (see Table 1). Preliminary analyses were conducted on the first plausible value only and thus sampling variance across plausible values has not been taken into account. The results combining the five analyses across plausible values will be conducted in October. The data from the mental health service are in the process of undergoing data quality checks and will be analyzed in November and December.

Conclusions:

Description of conclusions and recommendations based on findings and overall study.

Because the analyses are not complete, we cannot make conclusions at this time about the findings of the study. Additionally, given the correlational nature of our study, we will be providing recommendations regarding possible experimental methods to investigate any of the promising findings that we do uncover within our research. To the extent that cognitive age actually does affect response option usage (such as an extremity bias), reliability of self-report can be compromised, in particular, for younger respondents. Experimental studies to evaluate such loss of reliability would be important to conduct and thus we hope to outline possible directions to test the hypotheses.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

- Borgers, N., de Leeuw, E., & Hox, J. J. (1999). Surveying children: Cognitive development and response quality in questionnaire research. In A. Christianson et al. (eds). *Official statistics in a changing world*, pp. 133-140. Stockholm, SE: Statistics Sweden.
- Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Methodologie Sociologique*, April, 66, 60-75.
- Borgers, N., Hox, J., & Sikkels, D. (2003). Response quality in survey research with children and adolescents: the effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research*, 15, 83-94.
- de Leeuw, E. (2005). Surveying children. In Best, S. G. (Ed.) *Polling America: An Encyclopedia of Public Opinion*. Westport, CT: Greenwood Press.
- de Leeuw, E., Borgers, N., & Smits, A. (2004). Pretesting questionnaires for children and adolescents. In Presser, S., Rothgeb, J. M., Couper, M. T., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (eds.) *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: John Wiley & Sons.
- Dillman, D. (2000). *Mail and telephone surveys: The tailored design method*. New York, NY: Wiley.
- Holaday, B., & Turner-Henson, A. (1989). Response effects in surveys with school-age children. *Nursing research*, 38, 248-250.
- Institute for Education Sciences (2009a). *An intervention to enhance expository text comprehension via text structure instruction for primary-grade at-risk students*. Retrieved September 30, 2009 from <http://ies.ed.gov/funding/grantsearch/details.asp?ID=388>.
- Institute for Education Sciences (2009b). *Embedding knowledge-focused reading comprehension strategies in cumulative content-area instruction in grades 3-4-5: An assessment of immediate, transfer, and long-term impact on reading achievement*. Retrieved September 30, 2009 from <http://ies.ed.gov/funding/grantsearch/details.asp?ID=183>
- National Center for Education Statistics, (2006). *User's guide for the progress in international reading literacy study (PIRLS)* Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Rebok, G., Riley, A., Forrest, C., Starfield, B., Green, B., Robertsoon, J., & Tambor, E. (2001). Elementary school-aged children's reports of their health: A cognitive interviewing study. *Quality of Life Research*, 10, 59-70.
- Stanford, E. A., Chambers, C. T., & Craig, K. D. (2006). The role of developmental factors in predicting young children's use of a self-report scale for pain. *Pain*, 120, 16-23.
- Stapleton, L. M., Cafarelli, M., Almario, M. J., & Ching, T. (in press). Prevalence and characteristics of student attitude surveys used in public elementary schools in the United States. *Practical Assessment, Research and Evaluation*.
- Stone W. L., & Lemanek, K. L. (1990). Developmental issues in children's self-reports, pp. 18-82. In La Greca, A. M. (Ed.), *Through the Eyes of the Child*. Needham Heights, MA: Allyn and Bacon.

- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. New York, NY: Cambridge University Press.
- Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2004). Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. *Research on Social Work Practice, 14*, 191-200.

Appendix B. Tables and Figures

Not included in page count.

Figure 1: Response frequencies for students in the lowest and highest quartiles to the question “Students in my school help each other with their work”

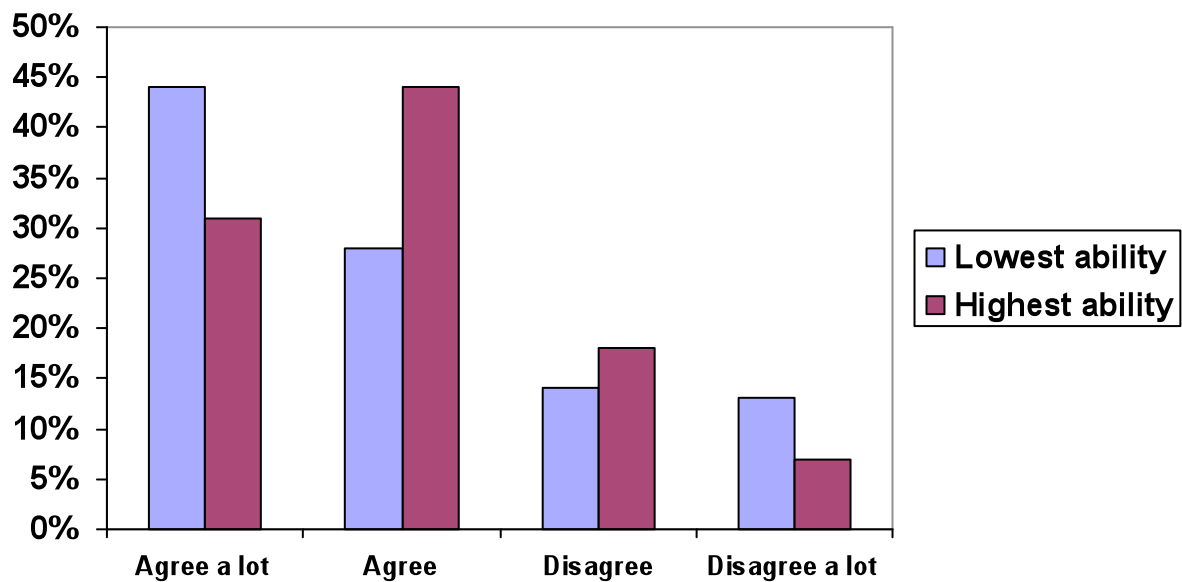


Table 1

Logit associated with 1st, 2nd, and 3rd thresholds with a one standard deviation change in Reading ability (First plausible value only)

Question	Parameter	Estimate	SE	p
I like being in school	1 st threshold	-.196	.03	<.001
	2 nd threshold	.028	.04	.428
	3 rd threshold	.196	.04	<.001
I feel safe when I am at school	1 st threshold	.183	.03	<.001
	2 nd threshold	.371	.05	<.001
	3 rd threshold	.516	.06	<.001
Students in my school show respect to each other	1 st threshold	-.087	.03	.009
	2 nd threshold	.092	.03	.008
	3 rd threshold	.360	.05	<.001
Students in my school care about each other	1 st threshold	-.058	.03	.102
	2 nd threshold	.157	.04	<.001
	3 rd threshold	.390	.05	<.001
Students in my school help each other with their work	1 st threshold	-.222	.03	<.001
	2 nd threshold	-.021	.03	.459
	3 rd threshold	.178	.04	<.001